Wavelets: Where Vision, Math & DSP Meet

– a quick introduction

Jackie (Jianhong) Shen

School of Math, University of Minnesota, Minneapolis

www.math.umn.edu/~jhshen

Knowledge learned from my former Ph.D. advisor

Gilbert Strang

Department of Math, MIT, Cambridge

www-math.mit.edu/~gs

Overview

- Seeking the simple codes of complex images
- Representation: learning from our own vision
- Image zooming, and zooming neurons
- Multiresolution framework of Mallat and Meyer
- Two key equations for Shape Function & Wavelet
- The fundamental theorem of Multiresolution
- 2-channel orthogonal & biorthogonal filter banks
- Applications
- (Chris and Michelle's talks are next)

Behind complexity is simplicity

Examples:

- The universal path to chaos is *period doubling*.
- (Biology) ACTG encode the complexity of life.
- (Computer) "0" and "1" (or spin up and down for *Quantum Computers*) are the digital "seeds."
- (Physics) The complexity of the material world is based on the limited number of basic particles.
- (Fractals) Simple algebraic rules hidden in complex shapes.

Conclusion:

Hidden in a complex phenomenon, is its simple evolutionary codes or building blocks (or the *atoms*).

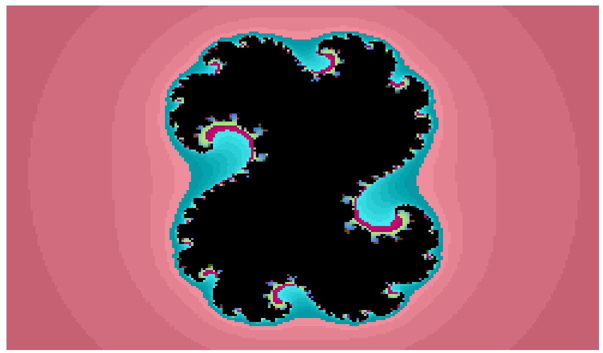
The complexity of image signals

<u>Images:</u>

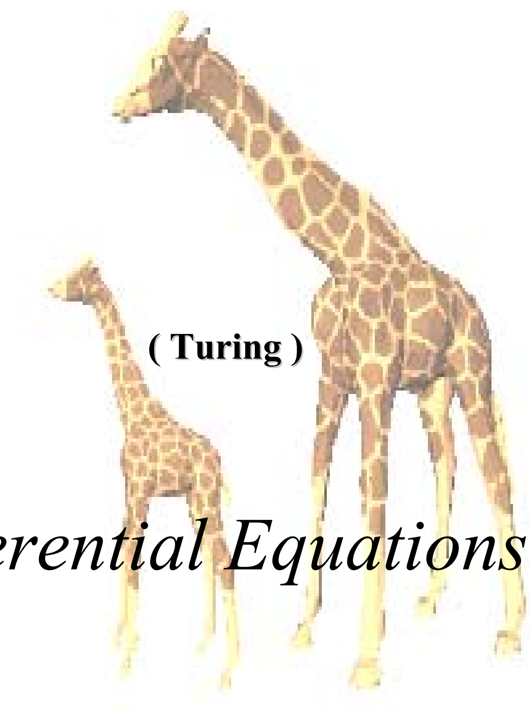
- Large dynamic range of scales.
- Often no good regularity as functions.
- Rich variations in intensity and color.
- Complex shapes and boundaries of "objects."
- Noisy or blurred (astronomical or medical image).
- "The lost dimension" --- range is lost but depth is still important for understanding meaningfully 2-D images.

Searching for the hidden code of images (I)

- Fractals: by *Iterated Function Systems*.



(Mandelbrot)



(Turing)

- Pattern formation: via *Differential Equations*.

Searching for the hidden code of images (II)

Statistical modeling (Geman's, Mumford, Zhu, Yuille…):

- – Image prior models (edge, regularity,...).
- – Image data models (noise, blurring,...).
- – Image disocclusion models.

- Parametric methods & lattice models.

- Non-parametric methods & learning via the maximum entropy principle.

A representation, not an interpretation...

- Benoit Mandelbrot (interview on France-Culture): "The world around us is very complicated. The tools at our disposal to *describe* it are very weak."

- Yves Meyer (1993):

  "Wavelets, whether they are …, will not help us to *explain scientific facts,* but they serve to *describe* the reality around us, whether or not it is scientific."

- Thus, to represent a signal, is to find a good way to *describe* it, not to *explain* the underlying physical process that generates it.

General images

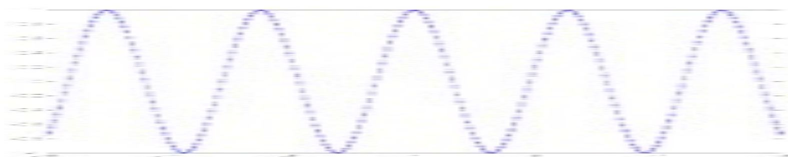
- Mostly no global multi-scale self-similarity.
- Contain both man-made and natural “objects”
- Mostly no simple and universal underlying physical or biological processes that generate the patterns in a general image.

Fourier was born too early…

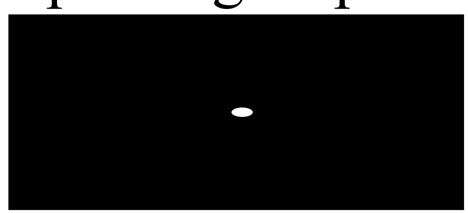
<u>Claim</u>: Harmonic waves are **bad** *vision neurons*…

Proof.

- A typical Fourier neuron is $\phi = \exp(iax)$.



- To "see" a simple bright spot $\delta(x)$ in the visual field,



all such neurons have to respond to it (!) since

$$\langle \delta , \phi \rangle \equiv 1 .$$

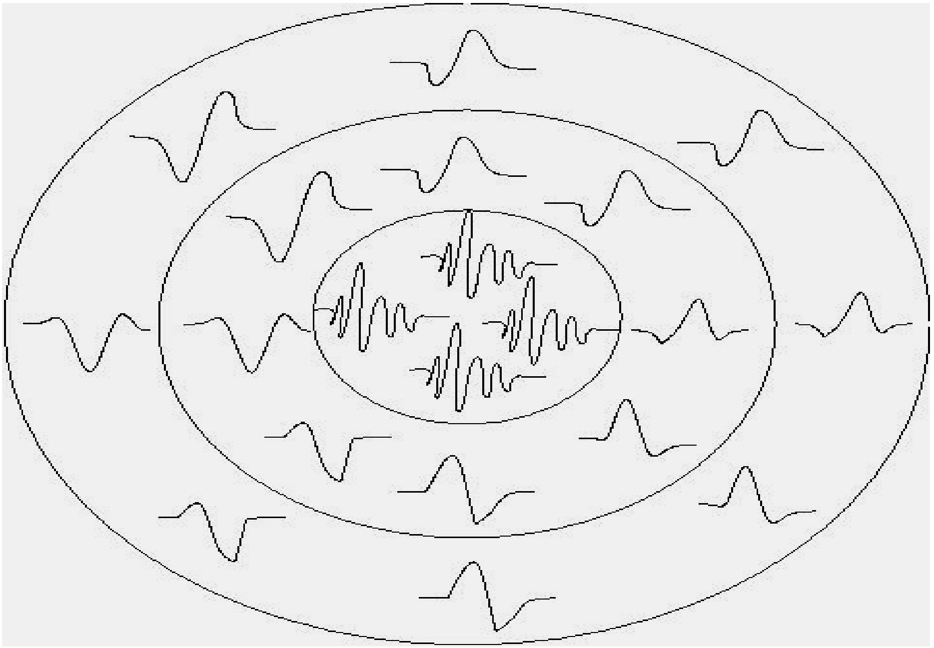
Efficiency of representation

- David Field (Cornell U, Vision psychologist):

  "To discriminate between objects, an effective transform (representation) encodes each image using the *smallest* possible number of neurons, chosen from a large pool."

Asking our own "headtop"…

- Psychologists show that visual neurons are *spatially* organized, and each behaves like a small sensor (receptor) that can respond strongly to spatial changes such as edge contours of objects (*Fields, 1990*).



The Marr's edge neuron model

- Detection of edge contours is a critical ability of human vision (Marr, 1982).

- Marr and Hildreth (1980) proposed a model for human detection of edges at all scales. This is Marr's *Theory of Zero-Crossings*:

$$G_\sigma = \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right),$$

$$\Psi_\sigma = \Delta G_\sigma = -\frac{2}{\sigma^2}\left(1-\frac{x^2+y^2}{2\sigma^2}\right)\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right),$$

Edge occurs in I where $(\Psi_\sigma * I) = 0.$

Haar's average-difference coding

- Marr's *edge detector* is to use second derivative to *locate* the maxima of the first derivative (which the edge contours pass through).

- *Haar Basis* (1909) encodes (modern language :-) the edges into image representation via the first derivative operator (i.e. moving difference):

$$(\dots x_{2n}, x_{2n+1}, \dots) \leftrightarrow \left(\dots a_n = \frac{x_{2n} + x_{2n+1}}{2}, d_n = \frac{x_{2n} - x_{2n+1}}{2}, \dots \right)$$

A good representation should *respect* edges

- Edge is so important a feature in image and vision analysis.
- A good image representation (or basis) should be capable of providing the edge information easily.
- Edge is a *local* feature. Local operators like differentiation must be incorporated into the representation, as in the coding by the Haar basis.
- Wavelets improve Haar, while respecting the above edge representation principle.

What to expect from a good representation?

- Mathematically *rigorous* (i.e. a clean and stable analysis and synthesis program exists. FT & IFT…).
- Having an independent *digital formulation*, and computationally *fast* (FFT, FWT…).
- Capturing the *characteristics* of the input signals, and thus many existing processing operators (e.g. image indexing, image searching …) are *directly* permitted on such representation.

Understanding images mathematically

- Let $\sum$ denote the collection of "all" images. What is the mathematical structure of Σ ? Suppose that $f \in \Sigma$ is captured by a camera. Then Σ should be invariant under
  - Euclidean motion of the camera:

$$f(x) \to f(Qx+a), \quad Q \in O(2), a \in R^2.$$

  - Flashing:

$$f(x) \to \mu f(x), \qquad \mu \in R^+,$$

or, more generally, a *morphological transform* ---

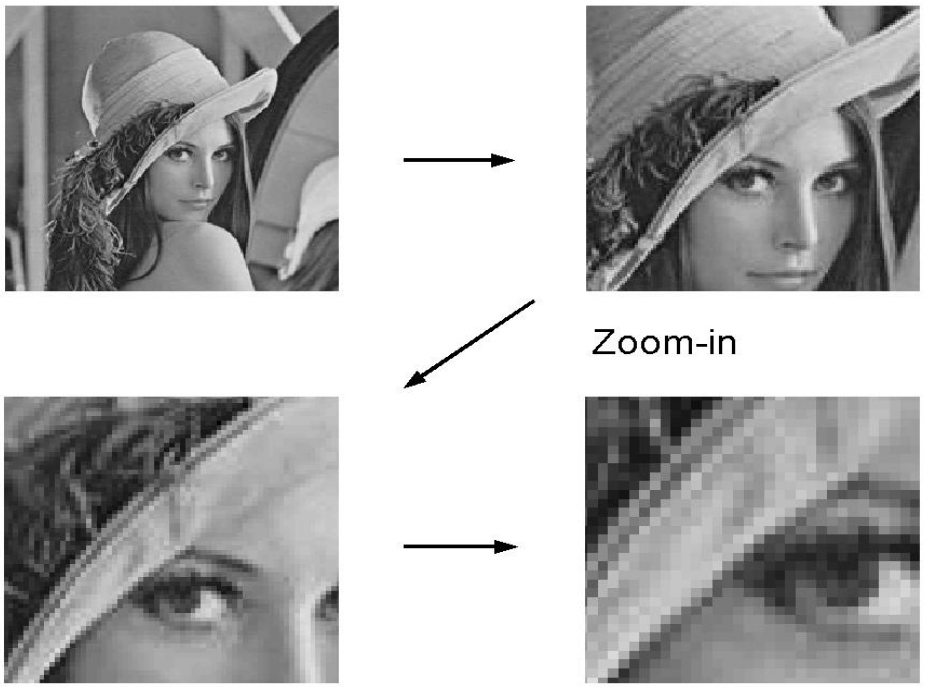
$$f(x) \to h(f(x)), \quad h: R \to R, h' > 0.$$

  - Zooming:

$$f(x) \to f(\lambda x), \quad \lambda \in R^+.$$

Let us focus on zooming

Zooming in 2-D



What is zooming?

- Zooming (aiming) center: a.
- Zooming scale: h.
- Zoom into the h-neighborhood at a in a given image I:

$$I_{a,h}(x) = I(a + h \cdot x), \quad x \in \Omega, \text{ the visual field;}$$

$$I_{a,h}\left(\tfrac{y-a}{h}\right) = I(y) \cdot 1_{\Omega}\left(\tfrac{y-a}{h}\right), \text{ the aperture.}$$

- Zooming is one of the most fundamental and characteristic operators for image analysis and visual communication. It reflects the ***multi-scale nature*** of images and vision.

The zooming neuron representation

- The zooming "**neuron**": $\psi(x)$.

- aiming (a) and zooming-in-or-out (h):

$$\psi_{a,h}(x) = \frac{1}{\sqrt{h}}\psi\left(\frac{x-a}{h}\right).$$

- Generating response (or neuron firing):

$$I_{a,h} = \langle I, \psi_{a,h} \rangle = \int I(x)\psi_{a,h}(x)\,dx\,.$$

- The zooming space: $(a, h) \in R \times R^{+}$.

A “good” neuron must be differentiating

- A *good* neuron should fire *strongly* to abrupt changes, and *weakly* to smooth domains (for purposes like efficient memory, object recognition, and so on).

- That means, for an uninteresting image I=c, the responses $I_{a,h}$ are all zeros:

  $$I_{a,h} = \langle I, \psi_{a,h} \rangle \equiv 0.$$

  This is the “**differentiating**” property of the neuron, just like “*d/dx*”:

  $$\int_R \psi(x) = 0.$$

The continuous wavelet representation

Definition:

A *differentiating* zooming neuron $\psi(x)$ is said to be a (continuous) *wavelet*. Representing a given image *I(x)* by *all* the neuron responses $I_{a,h} = \langle I, \psi_{a,h} \rangle$ is the corresponding *wavelet representation*.

Questions:

- Does there exist a “best” wavelet $\psi(x)$?
- Does a wavelet representation allow **perfect reconstruction**?

Synthesizing a wavelet representation

- Goal: to recover *perfectly* an image signal I from its wavelet representation $I(a,h)$.
- (Continuous) Wavelet synthesis:

$$I(a,h) = \left\langle I, \psi_{a,h} \right\rangle = \left\langle \hat{I}, \hat{\psi}_{a,h} \right\rangle = \left\langle \hat{I}\,\overline{\hat{\psi}}(h\xi),\, e^{-ia\xi} \right\rangle,$$

which is in the form of IFT. Thus $J(\xi,h) = \hat{I}\,\overline{\hat{\psi}}(h\xi)$ can be perfectly recovered via the a-FT of $I(a,h)$.

Then $\hat{I}$ can be perfectly recovered from J via

$$\hat{I}(\xi) = \int_{[0,\infty)} J(\xi,h)\,\hat{\psi}(h\xi)\,/\,h\;dh, \quad \int_0^\infty |\hat{\psi}(h)|^2\,/\,h\;dh = 1.$$

The admissibility condition & differentiation

- The *admissibility condition* of a continuous wavelet:

$$\int_0^\infty |\hat{\psi}(h)|^2 / h \, dh < \infty.$$

- A differentiating zooming neuron satisfies the AC since:

$$\hat{\psi}(0) = \int_R \psi(x)dx = 0, \text{ and } \hat{\psi}(h) = ch + o(h).$$

- Examples:
  - The Marr wavelet (Mexican-hat): second derivative of Gaussian.
  - The Shannon wavelet: $\psi(x) = 2\mathrm{sinc}(2x) - \mathrm{sinc}(x).$

The discrete set of zooming neurons

- Make a log-linear discretization to the scale parameter h:

$$j \to j = -\log_2 h_j = 0,\ \pm 1,\ \pm 2,\ \cdots .$$

- Make a *scale-adaptive* discretization of the zooming centers:

$$\text{at scale}\quad h_j = 2^{-j} : k \to a_k = kh_j = k/2^j,$$
$$k = 0, \pm 1, \pm 2, \cdots .$$

- The discrete set of zooming neurons:

$$\psi_{j,k}(x) = \frac{1}{\sqrt{h_j}}\psi\left(\frac{x-kh_j}{h_j}\right) = 2^{j/2}\psi(2^j x - k).$$

The discrete wavelet representation

- The wavelet coefficients:

$$d_{j,k} = \left\langle I, \psi_{j,k} \right\rangle = 2^{j/2} \int_R I(x) \psi(2^j x - k) dx.$$

$$d_{j,k} = I_{2^{-j}k, 2^{-j}}, \text{ in terms of the continuous WT.}$$

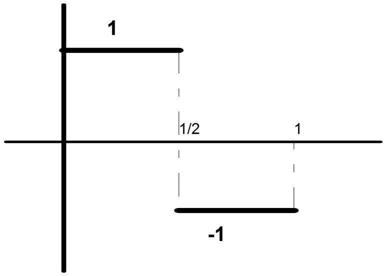
- <u>Questions:</u>
  - Does the set of all wavelet coefficients still encode the *complete* information of each input image I? Or equivalently,
  - Is the set of wavelets $\{\psi_{j,k}(x) : j, k \in Z\}$ a basis?

We don't know. But let's check out some examples...

Example 1: Haar wavelet

- The Haar "aperture" function is

$$\boxed{\psi^{\text{harr}}(x) = 1_{0 \le x < 1/2}(x) - 1_{1/2 \le x < 1}(x).}$$



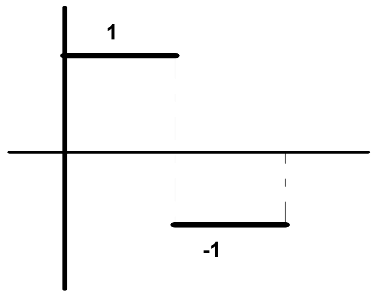
- Haar's theorem (1905):

All Haar wavelets $\psi_{j,k}^{\text{haar}}$ *, together with the constant function 1, consist into an* ***orthonormal basis*** *for the Hilbert space of all square integrable functions on* [*0, 1*].

Haar wavelets (cont'd)

- Haar's mother wavelet:

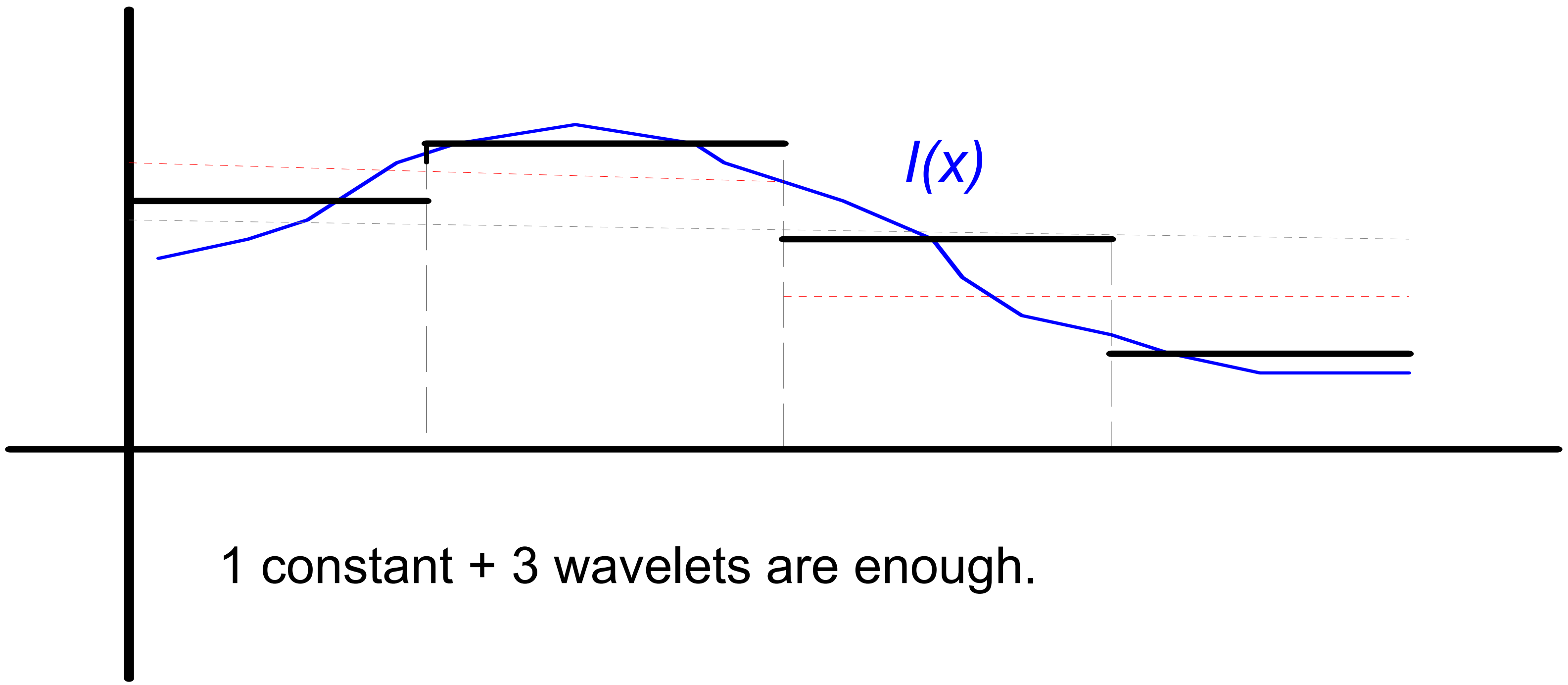
$$\boxed{\psi^{\text{Harr}}(x) = 1_{0\le x<1/2}(x) - 1_{1/2\le x<1}(x).}$$



- Why orthonormal basis?
  – Orthonormality is easy to see.
  – Completeness is due to the fact that:

    <u>All dyadically piecewise constant functions are dense in $L_2(0,1)$.</u>

Haar wavelets (cont'd)

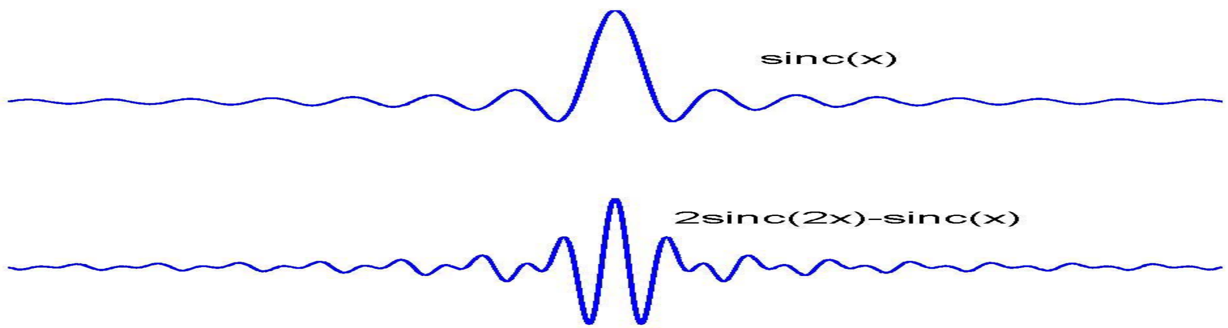


- Three Haar wavelets and the mean (constant) encode *all* the information of the piecewise constant approximation (or, the analog-to-digital transition).

Example 2: The Shannon wavelets

- The Shannon's "aperture" function is:

$$\psi^{\text{Shannon}}(x) = 2\,\mathrm{sinc}\,(2x) - \mathrm{sinc}\,(x).$$



- <u>Theorem</u>:

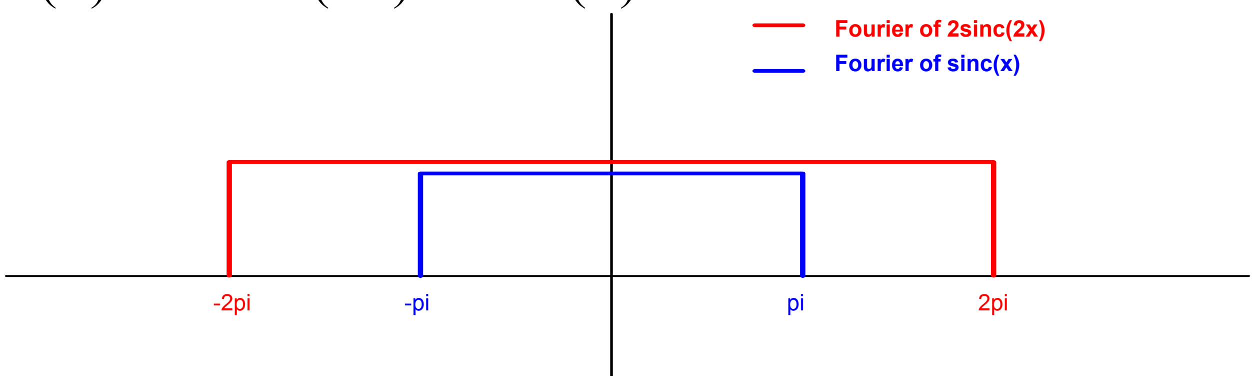
$\{\psi_{j,k}^{\text{Shannon}}(x) : j, k \in Z\}$ is an orthonormal basis of $L_2(\mathrm{R})$.

Shannon wavelets (cont'd)

- How to visualize the orthonormal basis ?

  Answer: go to the Fourier domain !

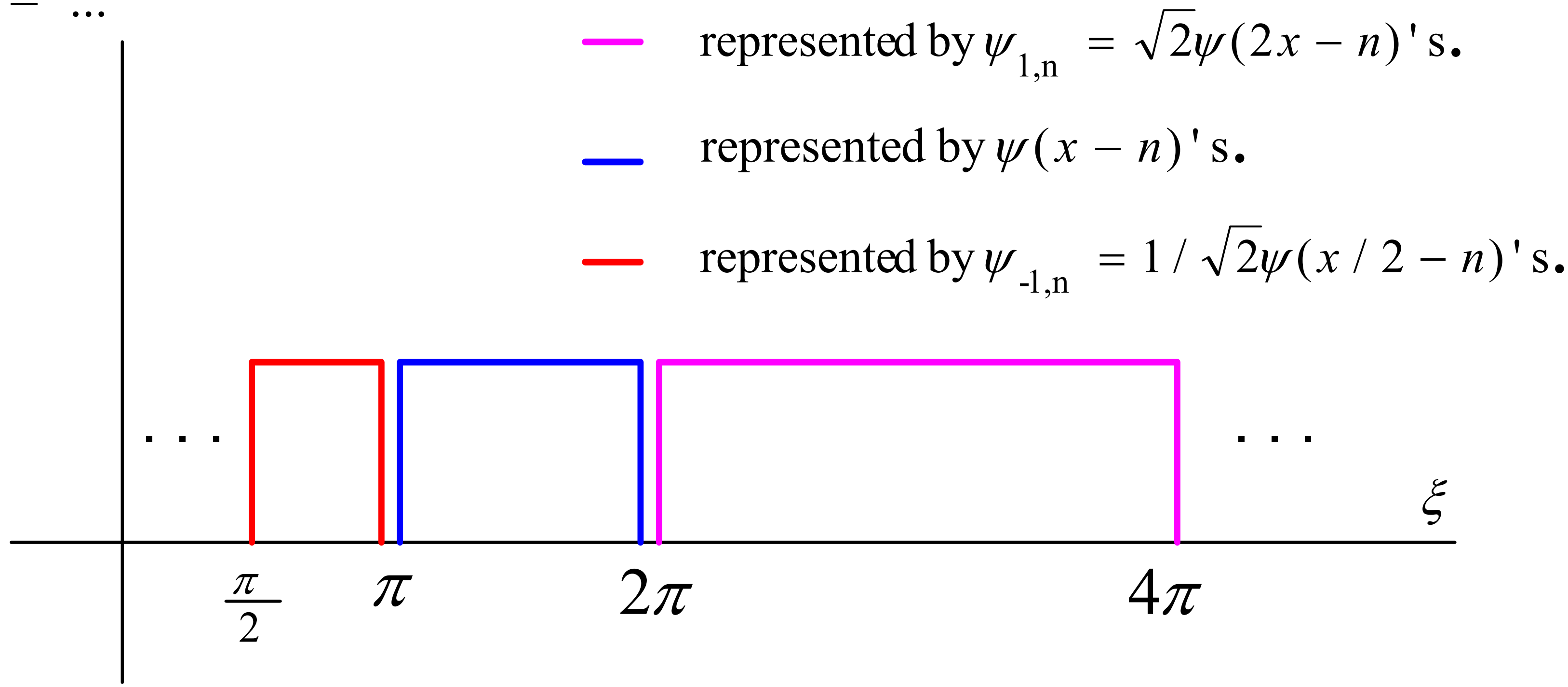
$$\psi^{\text{shannon}}(x) = 2\,\mathrm{sinc}\,(2x) - \mathrm{sinc}\,(x).$$



- According to Shannon:
  - All signals bandlimited to $(-\pi, \pi)$ can be represented by sinc(***x-n***)…
  - those bandlimited to $(-2\pi, \pi) \cup (\pi, 2\pi)$, by $\psi(x-n)$.
  - those bandlimited to $(-4\pi, 2\pi) \cup (2\pi, 4\pi)$, by $\psi_{1,n} = \sqrt{2}\psi(2x-n)$.
  - ...

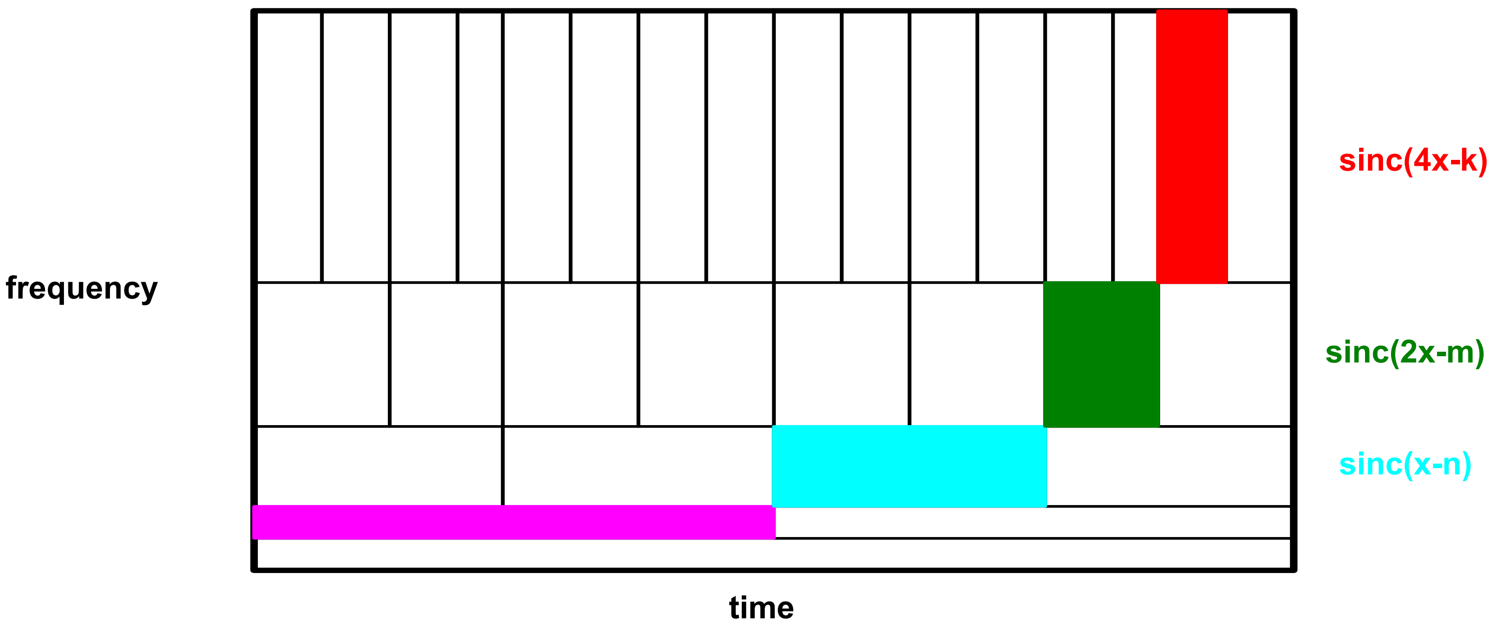
Shannon wavelets (cont'd)

- According to Shannon:
  - All signals bandlimited to ($-\pi$, π) can be represented by sinc($\boldsymbol{x}$-$\boldsymbol{n}$)…
  - those bandlimited to (-2π, $-\pi$) U (π, 2π), by $\psi(x-n)$.
  - those bandlimited to (-4π,-2π)U(2π, 4π), by $\psi_{1,n}=\sqrt{2}\psi(2x-n)$.
  - ...



Partition of the time-frequency plane

- Heisenberg's uncertainty principle requires that each TF atom must have: $\Delta t \cdot \Delta x \geq 2\pi$.
- Thus, for an *optimal* localization, the "life time" of an atom must influence its scale or frequency content.



Multiresolution analysis

Mallat and Meyer (1986):

An (orthogonal) multiresolution of $L_2(R)$ is a chain of closed subspaces indexed by all integers:

$$\cdots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \cdots$$

subject to the following three conditions:

- (**completeness**) $\overline{\lim_{n\to\infty} V_n} = L_2(R), \qquad \lim_{n\to-\infty} V_n = \{0\}.$
- (**scale similarity**) $f(x) \in V_n \Leftrightarrow f(2x) \in V_{n+1}.$
- (**translation seed**) V_0 has an orthonormal basis consisting of all integral translates of a single function $\phi(x) : \{\phi(x-n) : n \in Z\}.$

Equations for designing MRA

- The refinement (dilation) equation for the "seed" function:

$$\boxed{\phi(x) = 2\sum_n h_n \phi(2x-n), \text{ for a suitable set of } h_n\text{'s.}}$$

This seed function is called: scaling function, shape fcn…

- Where is the wavelet?

Let W_0 denote the orthogonal complement of V_0 in V_1. Then W_0 is also orthogonally spanned by the integer translates of a single translation seed $\psi(x)$, the **wavelet**!

$$\boxed{\psi(x) = 2\sum_n g_n \phi(2x-n), \text{ for a suitable set of } g_n\text{'s.}}$$

Wavelets representation

Theorem:

$\{\psi_{j,k} = 2^{j/2}\psi(2^j x - k) : j,k \in Z\}$ is an orthonormal basis for L_2.

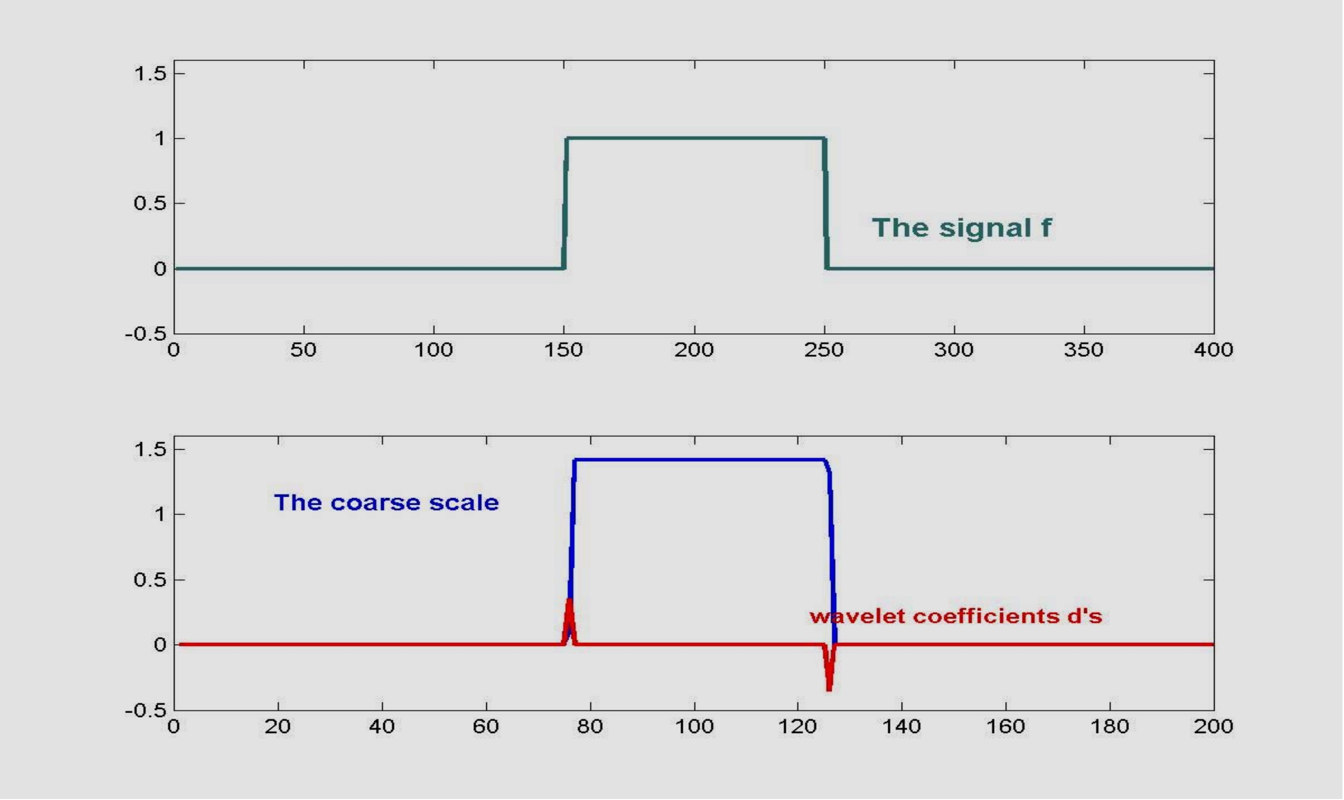
Wavelets representation of a signal:

$$\begin{array}{ccccccc} I \sim I_j \in V_j & \longrightarrow & I_{j-1} \in V_{j-1} & \longrightarrow & & \longrightarrow & I_0 \in V_{0.} \\ & \searrow + & & \searrow + & & \searrow + & \\ & & d_{j-1} \in W_{j-1} & & d_{j-2} \cdots & & d_0 \in W_0. \end{array}$$

$$I_j = d_{j-1} + d_{j-2} + \cdots + d_0 + I_0 .$$

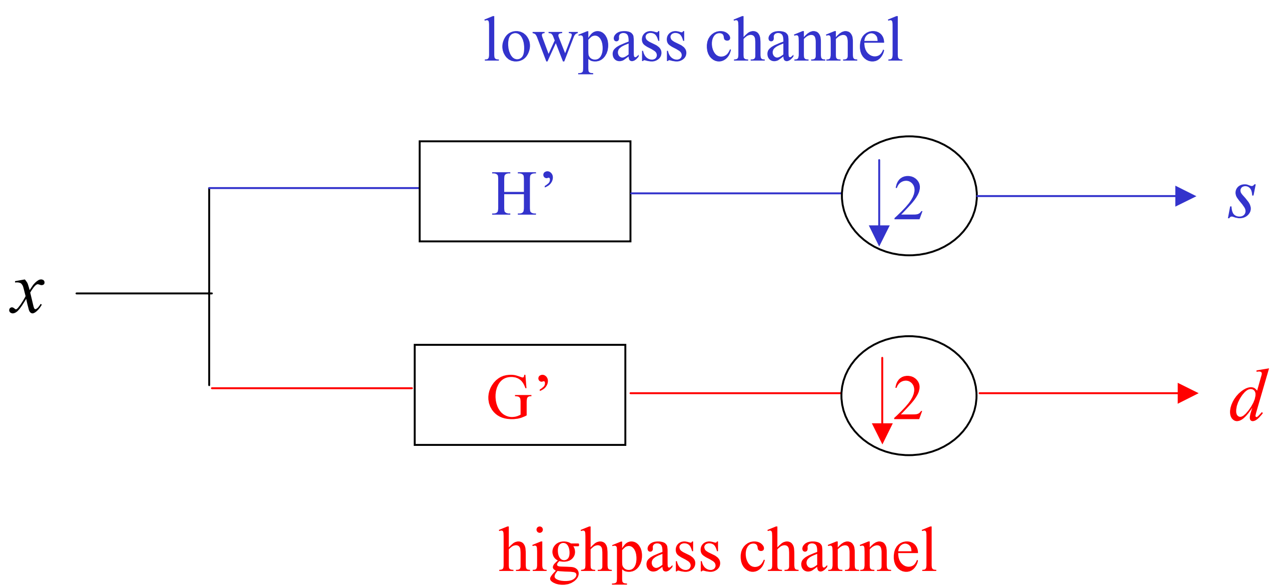
An example of wavelet decomposition

One level wavelet decomposition of a 1-D signal



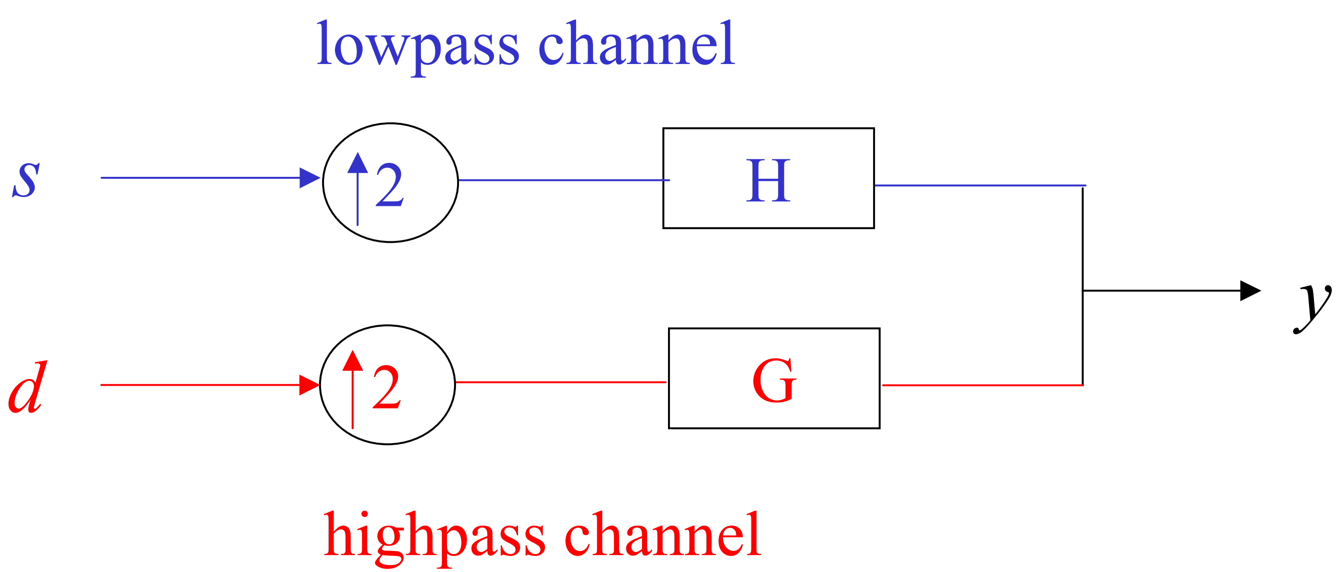
2-channel filter bank: Analysis bank

- H' is the <u>lowpass</u> filter and G' is the highpass filter.
- $\downarrow 2$ is the <u>downsampling</u> operator: (1 3 4 6 5) $\longrightarrow$ (1 4 5).

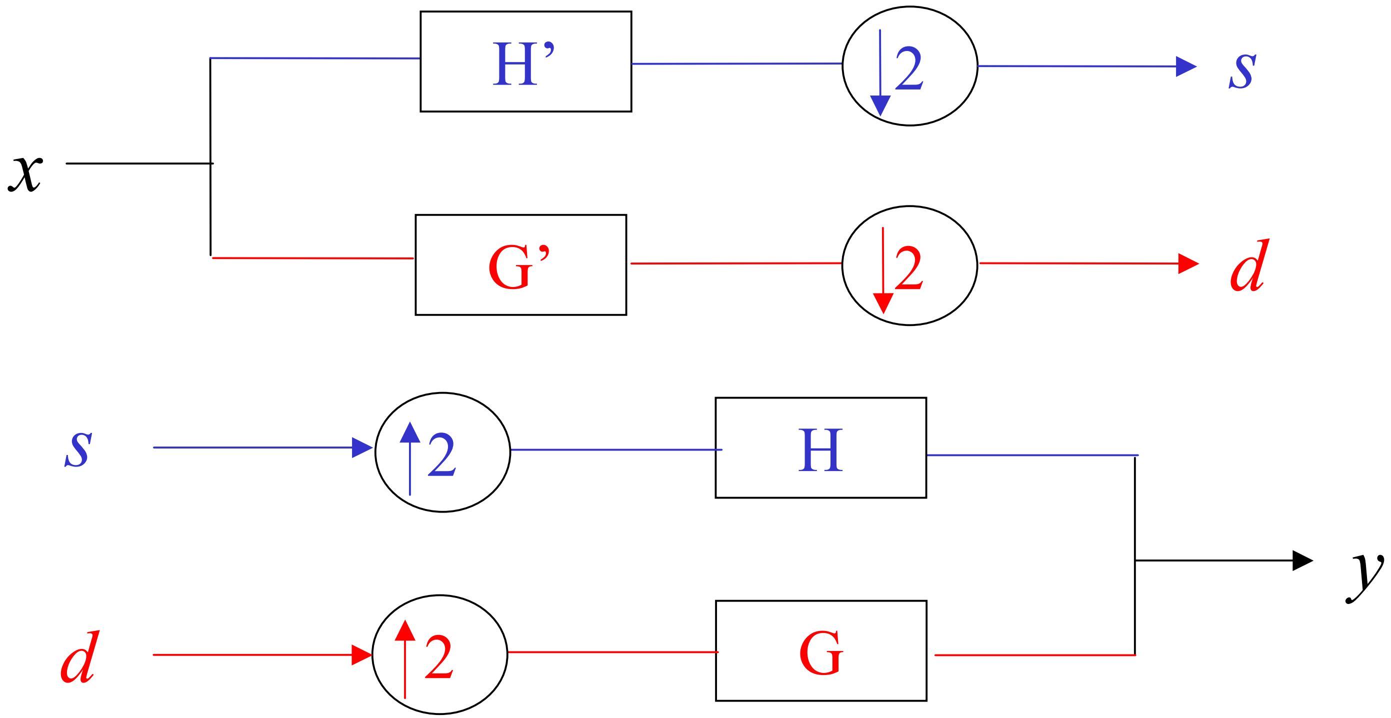


2-channel filter bank: Synthesis bank

- H is the <u>lowpass</u> filter and G is the highpass filter.
- $\uparrow 2$ is the <u>upsampling</u> operator: (1 4 5) $\longrightarrow$ (1 0 4 0 5).

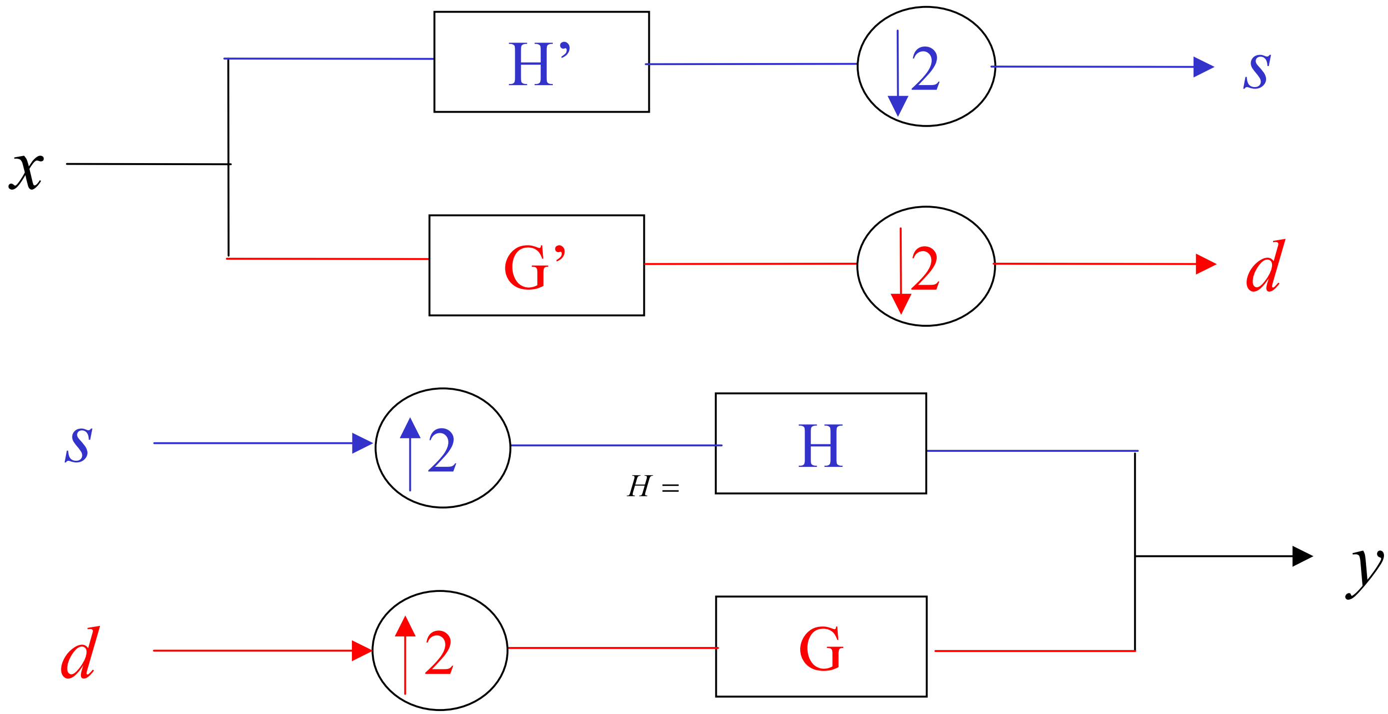


A *biorthogonal* filter bank



Biorthogonal (or perfect) filter bank: if $y=x$ for all inputs x .

An *orthogonal* filter bank



<u>Orthogonal filter bank</u>: if it is biorthogonal, and both *analysis* filters H' and G' are the *time reversals* of the *synthesis* filters H & G: H=(1, 2, 3) $\longrightarrow$ H'=(3, 2, 1).

The fundamental theorem of MRA

- An *orthogonal* Mallat-Meyer MRA corresponds to an *orthogonal* filter bank with the synthesis filters:

$$H = (h_n : n \in Z), \quad G = (g_n : n \in Z).$$

where, the h's and g's are the 2-scale *connection coefficients* in the dialation and wavelet equations:

$$\phi(x) = 2\sum_n h_n \phi(2x - n), \quad \psi(x) = 2\sum_n g_n \phi(2x - n).$$

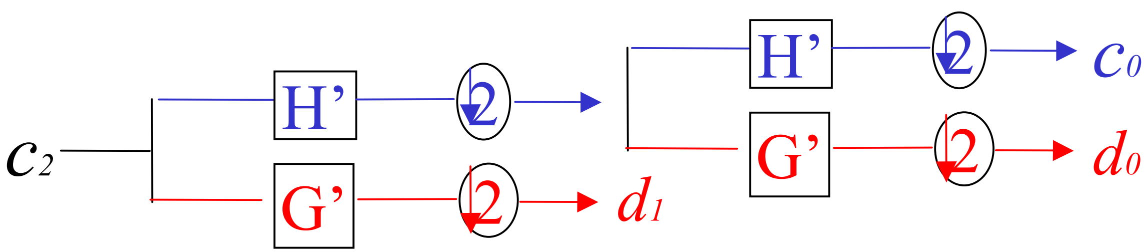
And, the *multiresolution* wavelet decomposition of f corresponds to the *iteration* of the analysis bank with the ϕ-coefficients of f as the input digital data.

The fundamental theorem (cont'd)

$$\begin{tikzcd} I_j \in V_j \arrow[r, dash] \arrow[dr, dash, "+"] & I_{j-1} \in V_{j-1} \arrow[r, dash] \arrow[dr, dash, "+"] & {} \arrow[r, dash] \arrow[dr, dash, "+"] & I_0 \in V_{0.} \\ & d_{j-1} \in W_{j-1} & d_{j-2} \cdots & d_0 \in W_0. \end{tikzcd}$$

$$I_j = d_{j-1} + d_{j-2} + \cdots + d_0 + I_0 .$$

Suppose j=2, and $I_2 = \sum_k c_2(k)\phi_{2,k}(x).$



Some major applications

- FBI fingerprints.
- JPEG2000.
- Image indexing and image search engines (for databank).
- Image modeling (such as MRF on the wavelets domain).
- Image denoising and restorations.
- Texture analysis.
- Direct processing tools on the wavelets domain.
- Algorithm speeding up based on multi-resolution rep..
- Time series analysis.
- A lot of others ...

New Directions of Wavelets

- Random Wavelets Expansion (RWE) by Mumford-Gidas **[2001]**, to model the scale-invariance of general images.

- Geometric Wavelets:
  - D. Donoho's school: ridgelets, wedgelets, curvelets.
  - S. Mallat [2001]: beamlets.
  - T. Chan & H.-M. Zhou [2000], A. Cohen [2002]: integrate computational PDE techniques such as the ENO scheme into wavelet transforms, to better capture shocks (discontinuities).

That is all, folks…

Thank you for your patience!